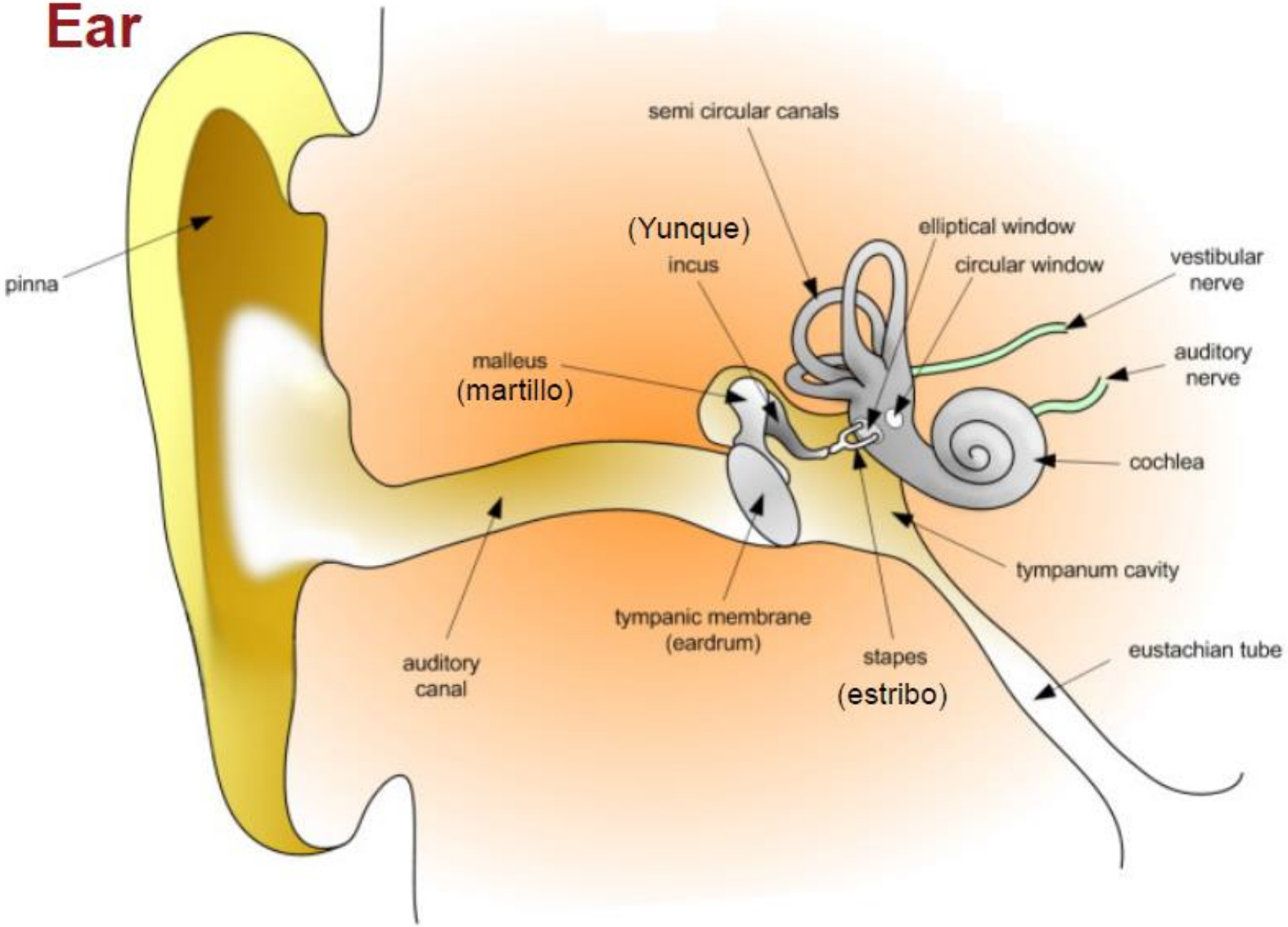
Hearing and Auditory Perception

Ear

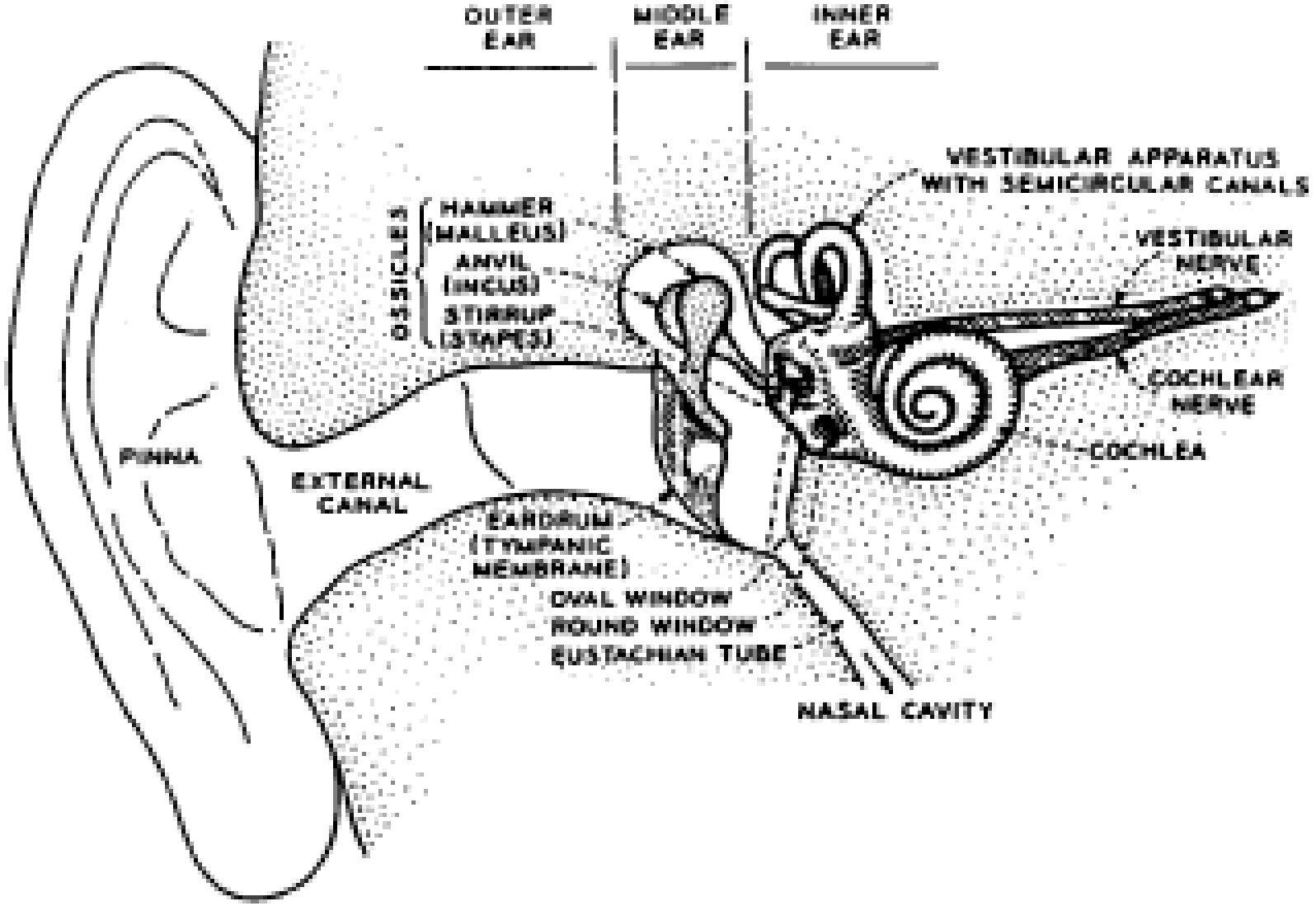


Anatomy of the ear

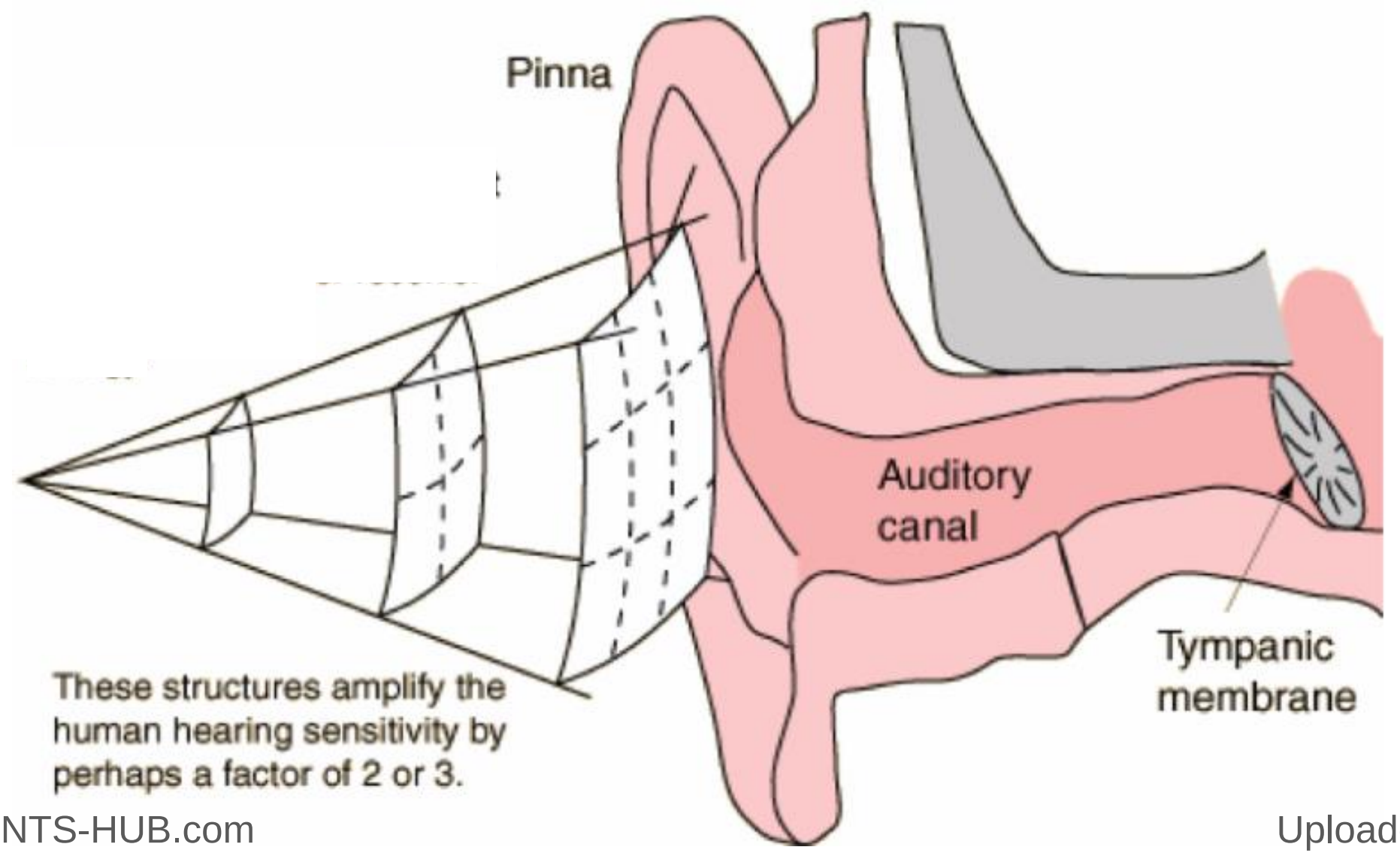
- The ear divided into three sections:
  - the outer
  - Middle
  - Inner ear

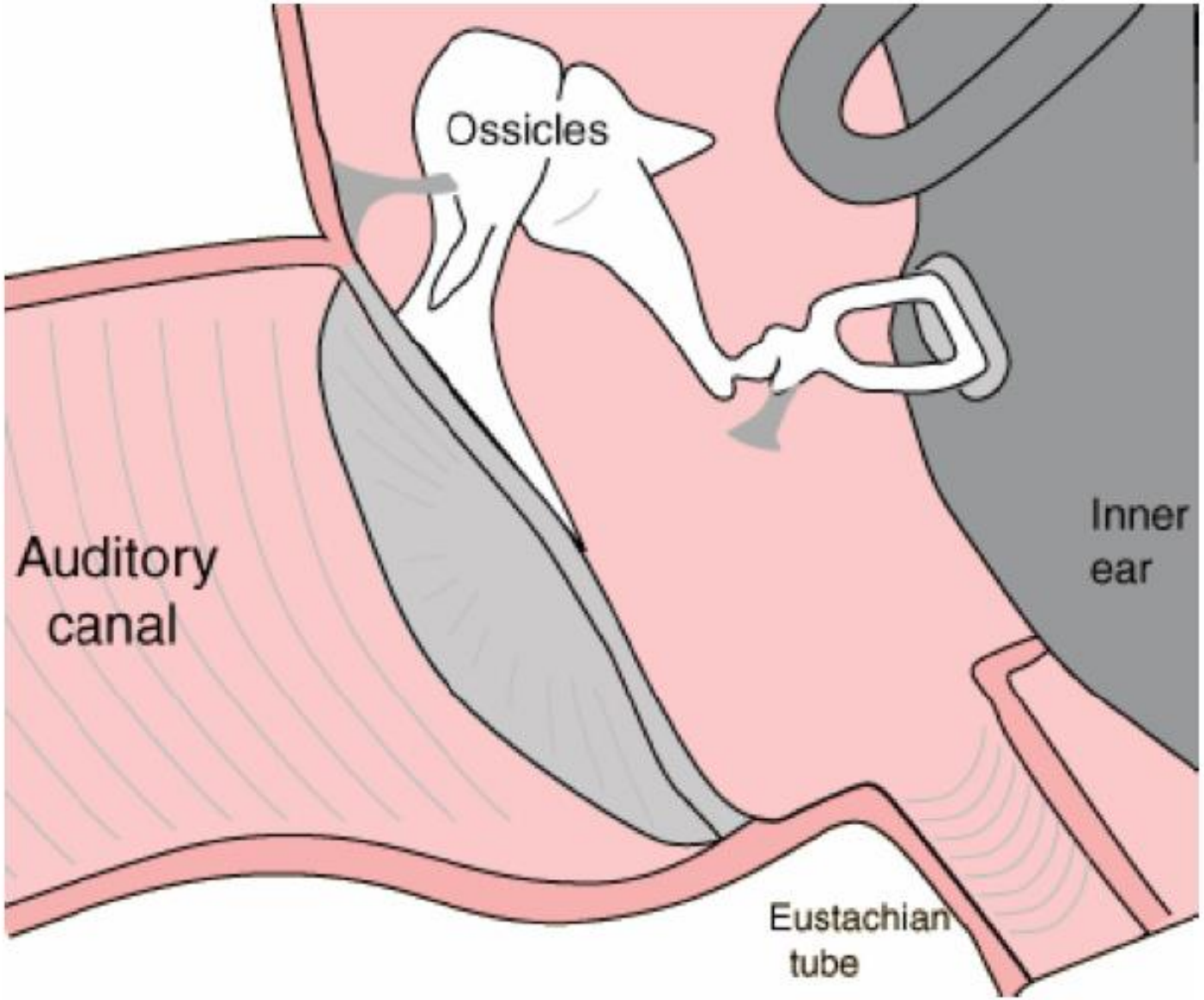
➢The outer ear is terminated by the eardrum (tympanic membrane).

➢Sound waves entering the auditory canal of the outer ear are directed into the ear drum and cause it vibrates.



The Outer Ear



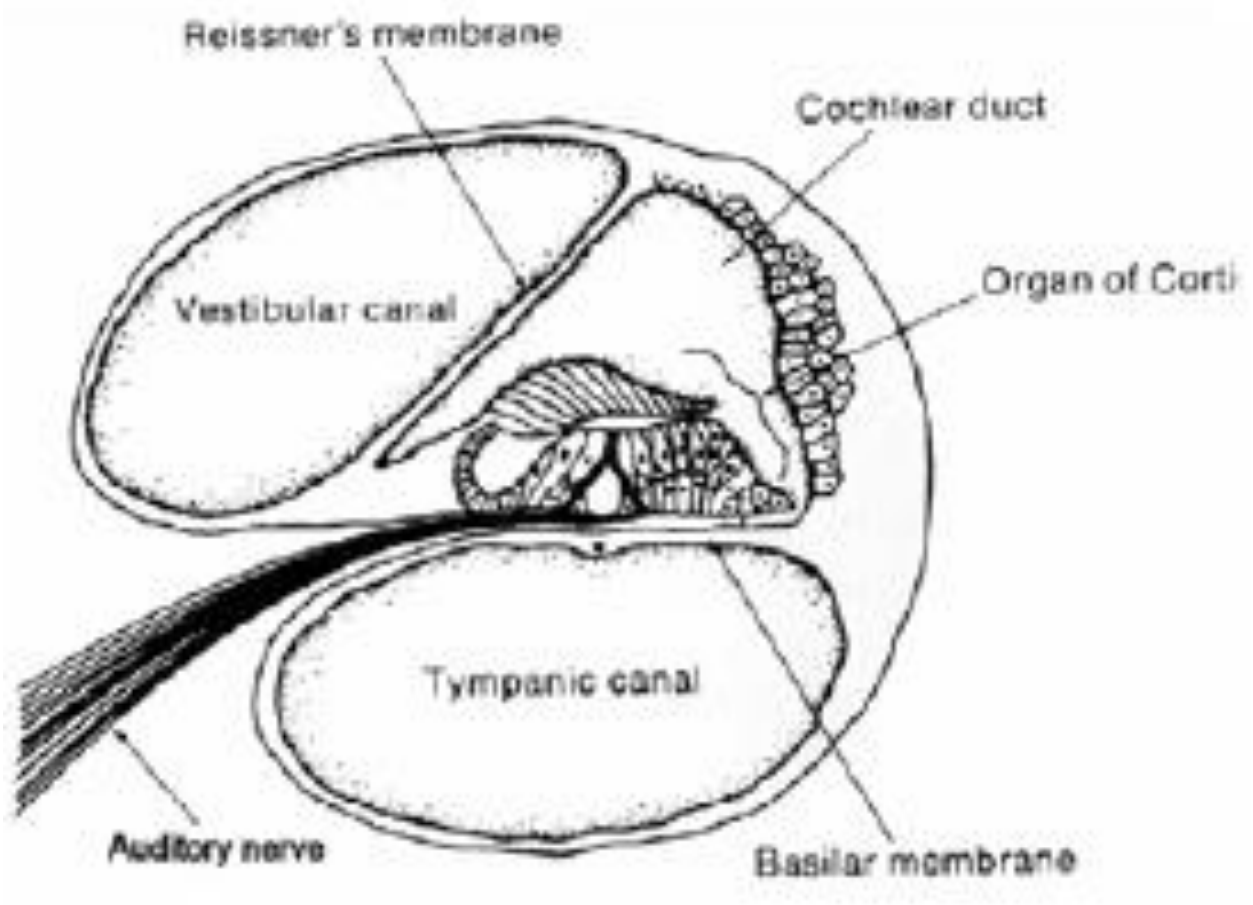


- The vibrations are transmitted by the middle ear, and air filled section comprising a system of three tiny bones, the malleus, incus, stapes, to the cochlea (the inner ear).

- The cochlea is a aspiral of about 23/4 turns which unrolled would be about 3.5cm long.

- The cochlea consists of three fluid-filled sections

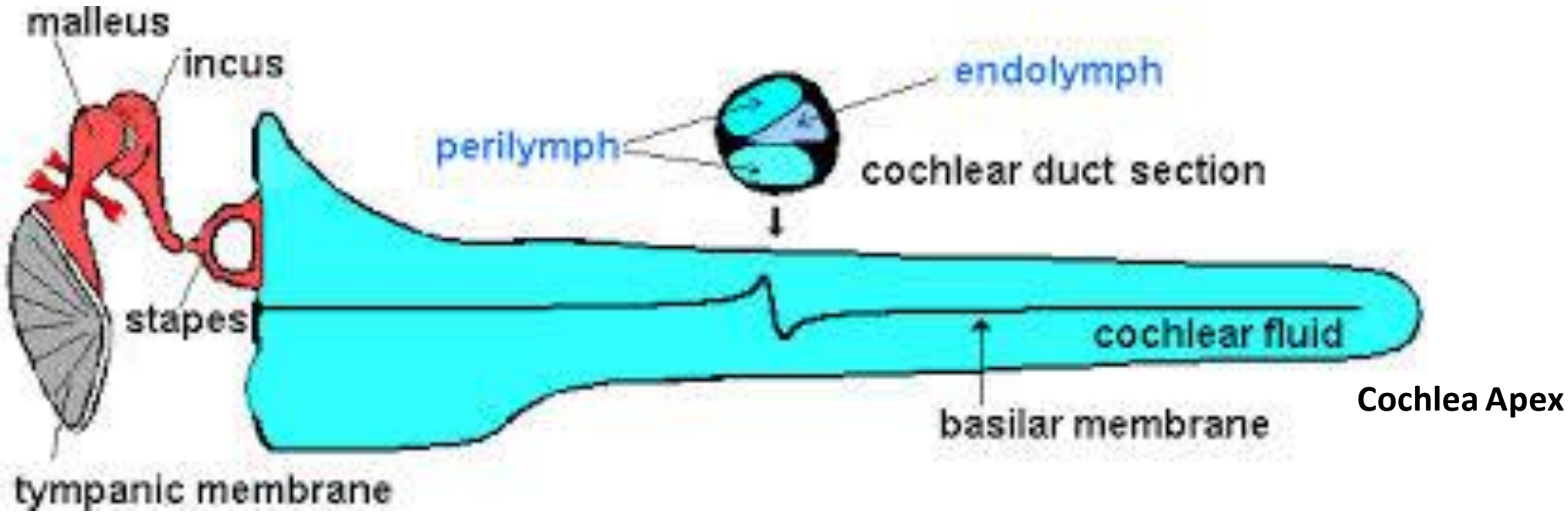
- One, the cochlea duct, is relativily small in cross-sectional area, and the other two, the scala vestibule and scala tympani are larger and roughly equal in area.

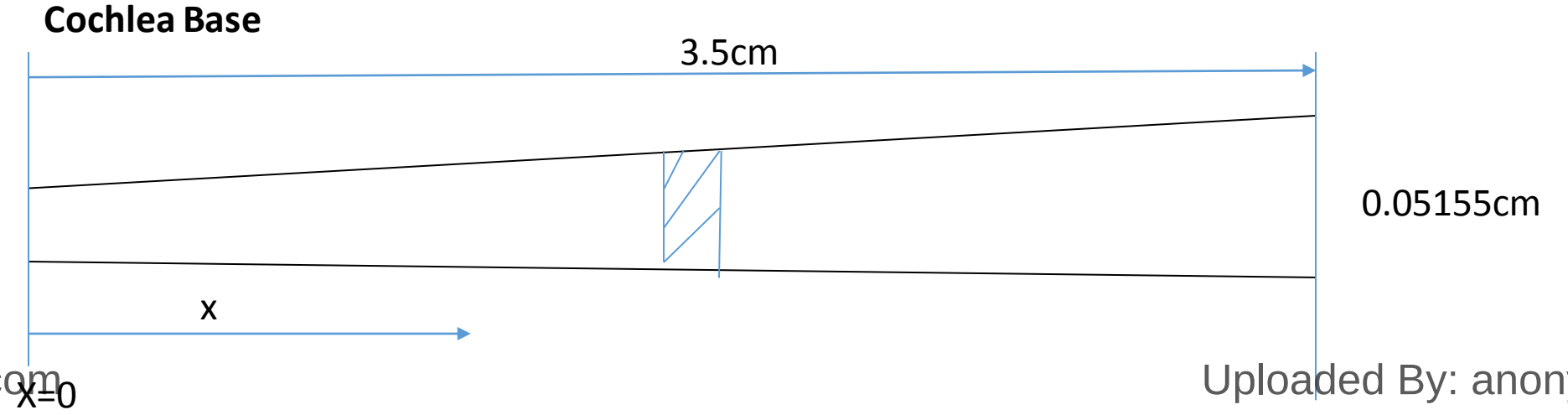


- The scala vestibule is connected to the stapes via aval window

- The scala tympani terminates in the round window which is a thin membranous cover allowing the free movement of the cochlear fluid

- Running the full length of the cochlea is the Basilar Membrane (BM) which separates the cochlea duct from the scala vestibule.

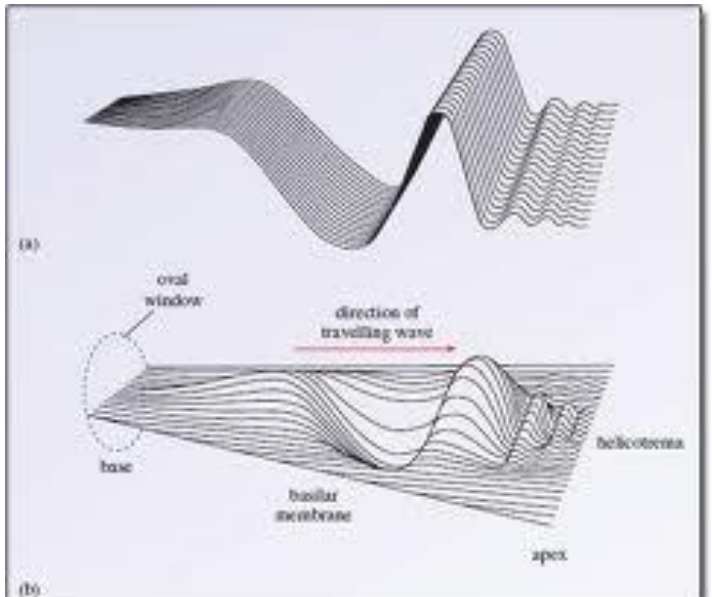




- It has been shown by Bekesey (1960) that when the vibrations of the eardrum are transmitted by the middle ear into movement of the stapes, the resulting pressure within the cochlea fluid generates a travelling wave of displacement on the basilar membrane.

- The location of the maximum amplitude of this travelling wave varies with frequency of the eardrum vibrations.

- The response of the BM at an instant of time to a pure tone at the stapes is schematically shown below

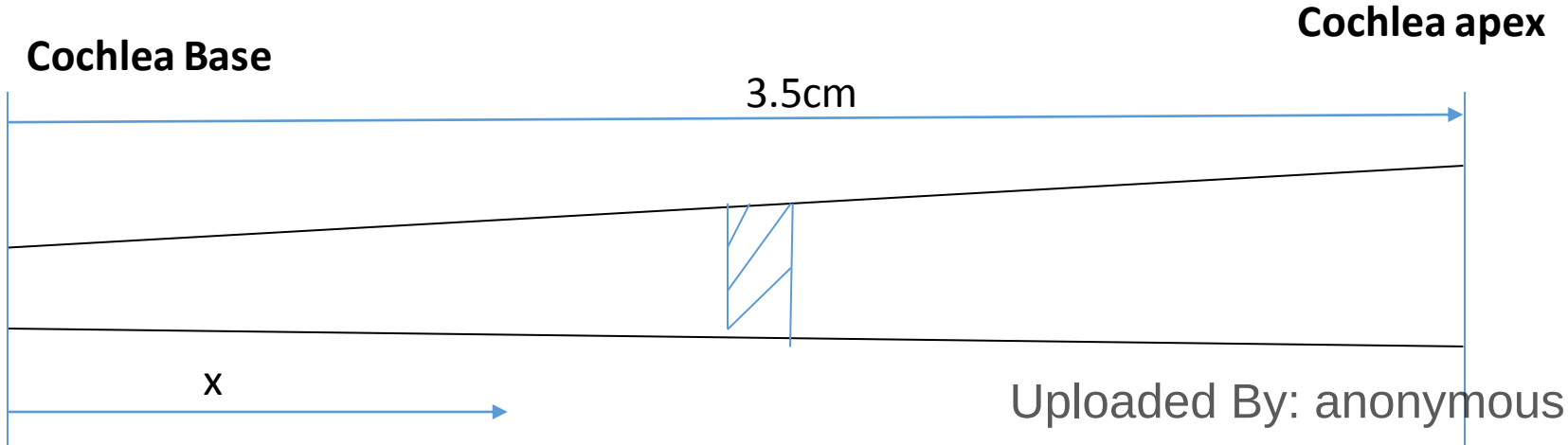


Video

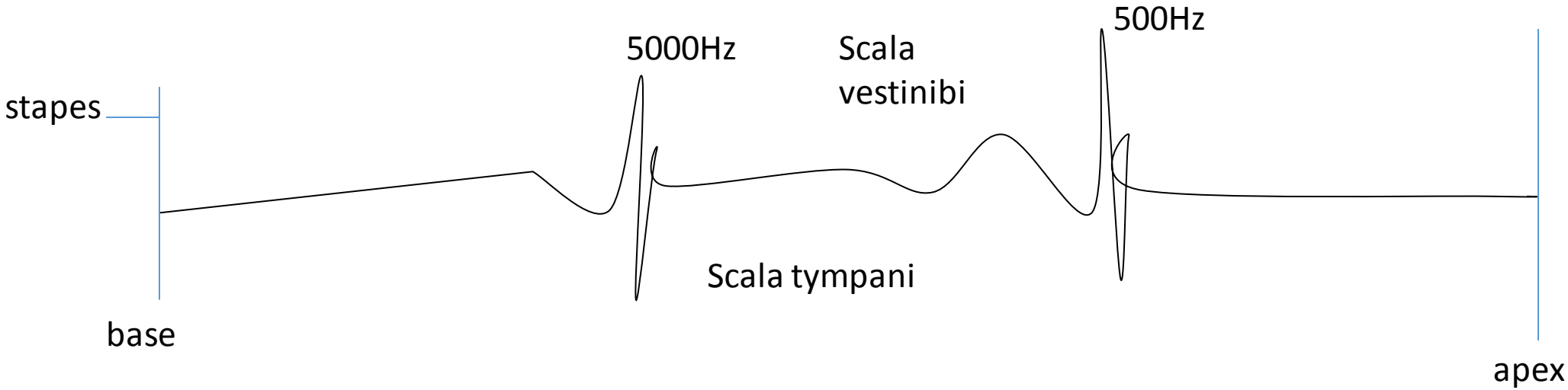
- The basilar membrane varies in width and stiffness along its length

- At the basal end it is narrow and stiff whereas towards the apes it is wider and more flexible.

- The maximum membrane displacement will occur at the stapes end for high frequencies and at the far end (apex) for low frequencies.



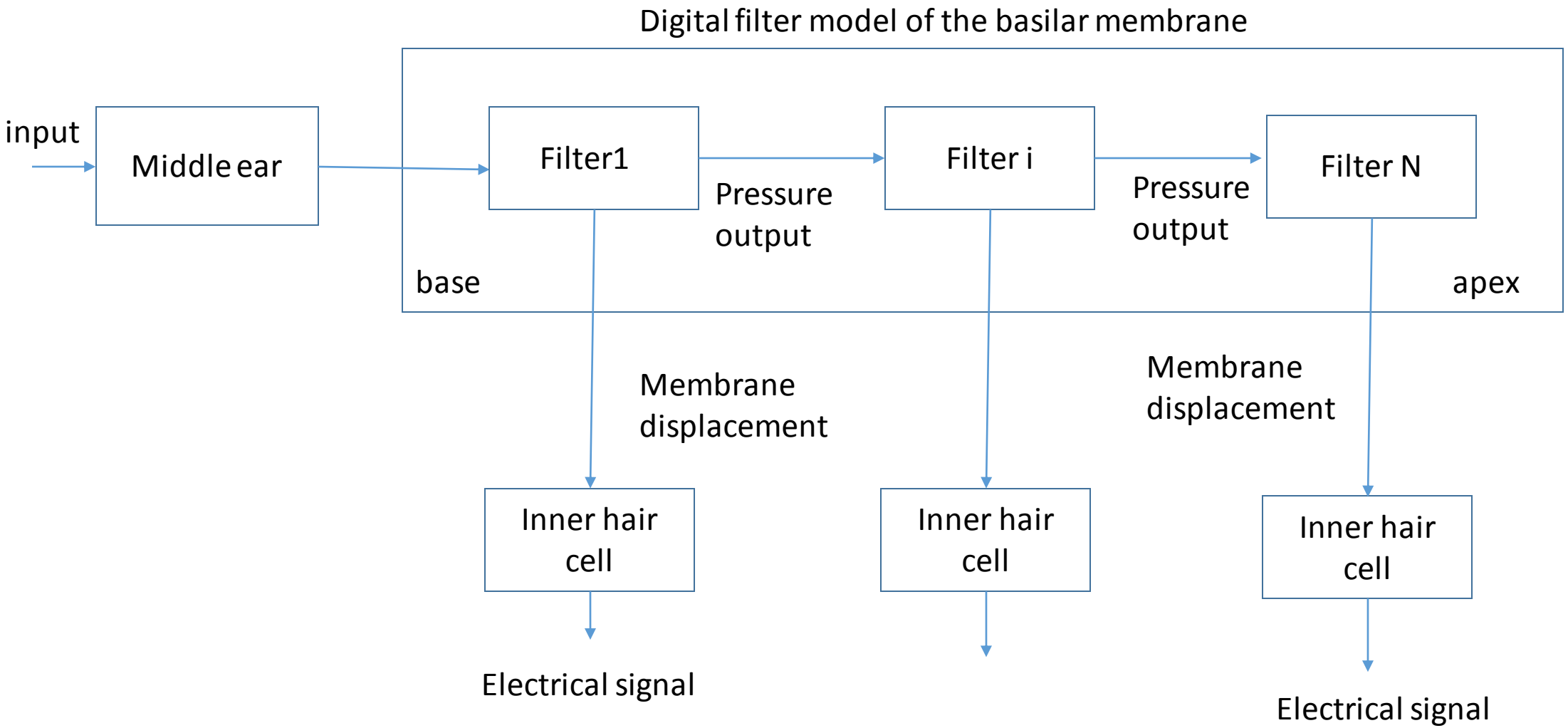
- The wave motion along the BM is governed by the mechanical prosperities of the membrane and hydrodynamic prosperities of the surrounding fluid (scalas)



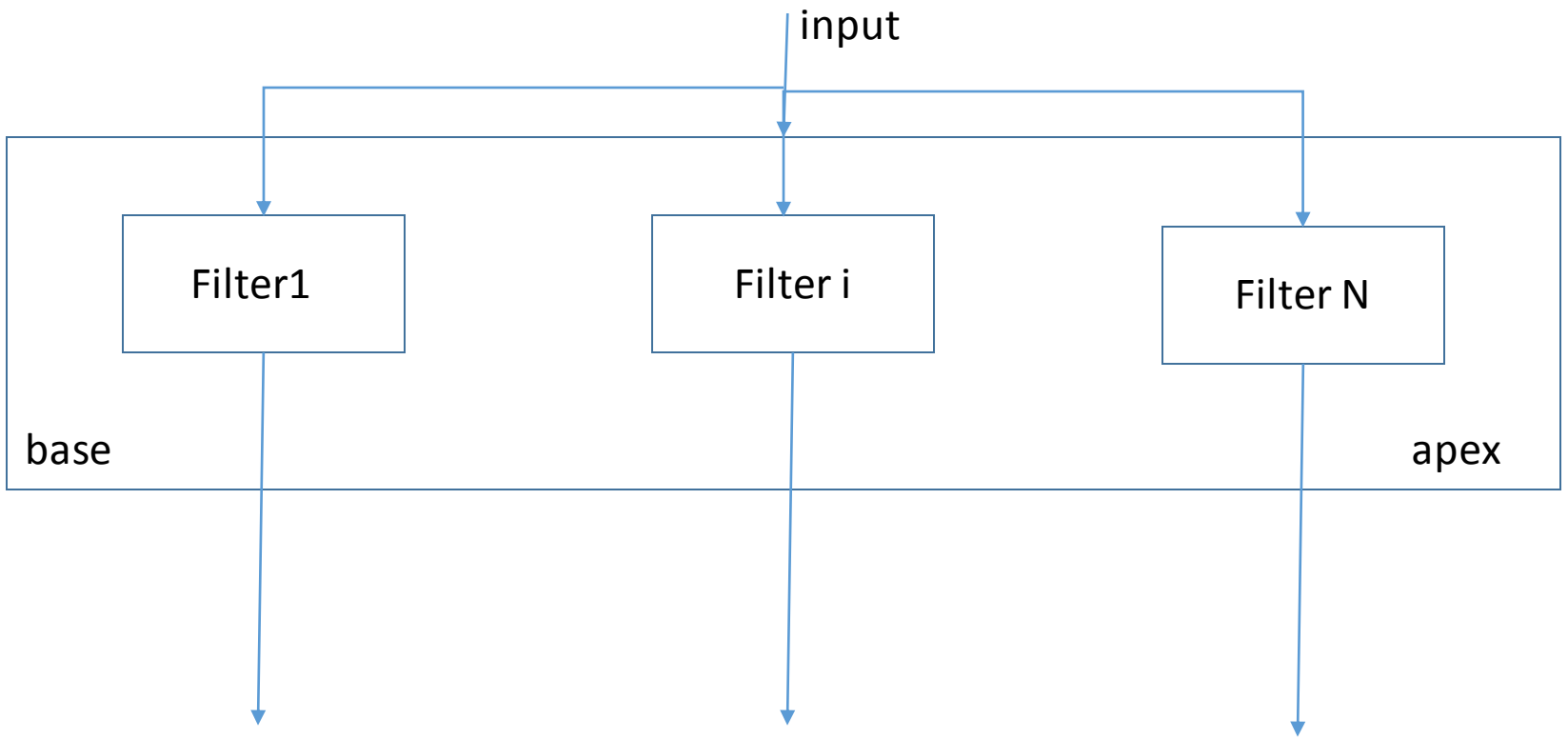
- It appears that each point of the BM moves independently (i.e. a point on the basilar membrane is assumed to have no direct mechanical coupling to neighboring points).

- However, the neighboring points are coupled through the surrounding fluid

Transmission Line Model



Parallel Filter Bank Model



Sound pressure level

- Atmospheric pressure is approximately 15 lb/in^2 or 1 bar. A variation of one millionth of the atmospheric pressure (or 1 micro bar) is an appropriate stimulus for hearing. Such a pressure variation is generated in normal conversation by the human voice.

- The minimal level of pressure changes to which man is sensitive is well over 0.0002 microbars.

- A figure commonly used at the upper limit of hearing is 2000 microbars.

- At this upper limit, acoustic stimulus is accompanied by pain. We know,

$$dB(power) = 10\log\frac{P_o}{P_i}$$

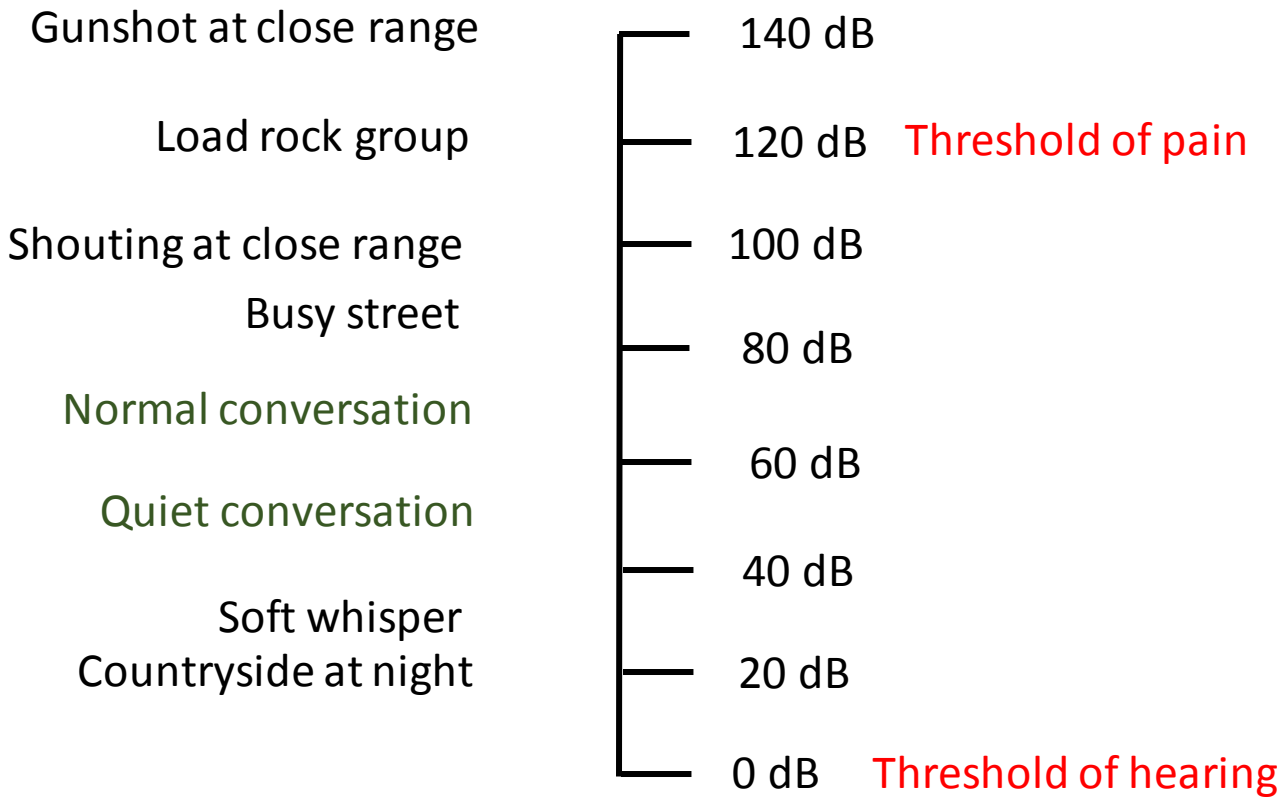
- Since acoustic power is directly related to the square of acoustic pressure,

$$dB(pressure) = 10\log[(P_o)^2/(p_o)^2] = 20\log[P_o/P_i]$$

Pi is commonly taken as 0.0002 micro bars (at or below the threshold for hearing)

- Given an upper limit P_o as 2000 microbars, the Sound Pressure Level (SPL)of an acoustic stimulus is:

$$SPL = 20\log(2000\,\mu bars/0.0002\,\mu bars) = 140\,dB$$

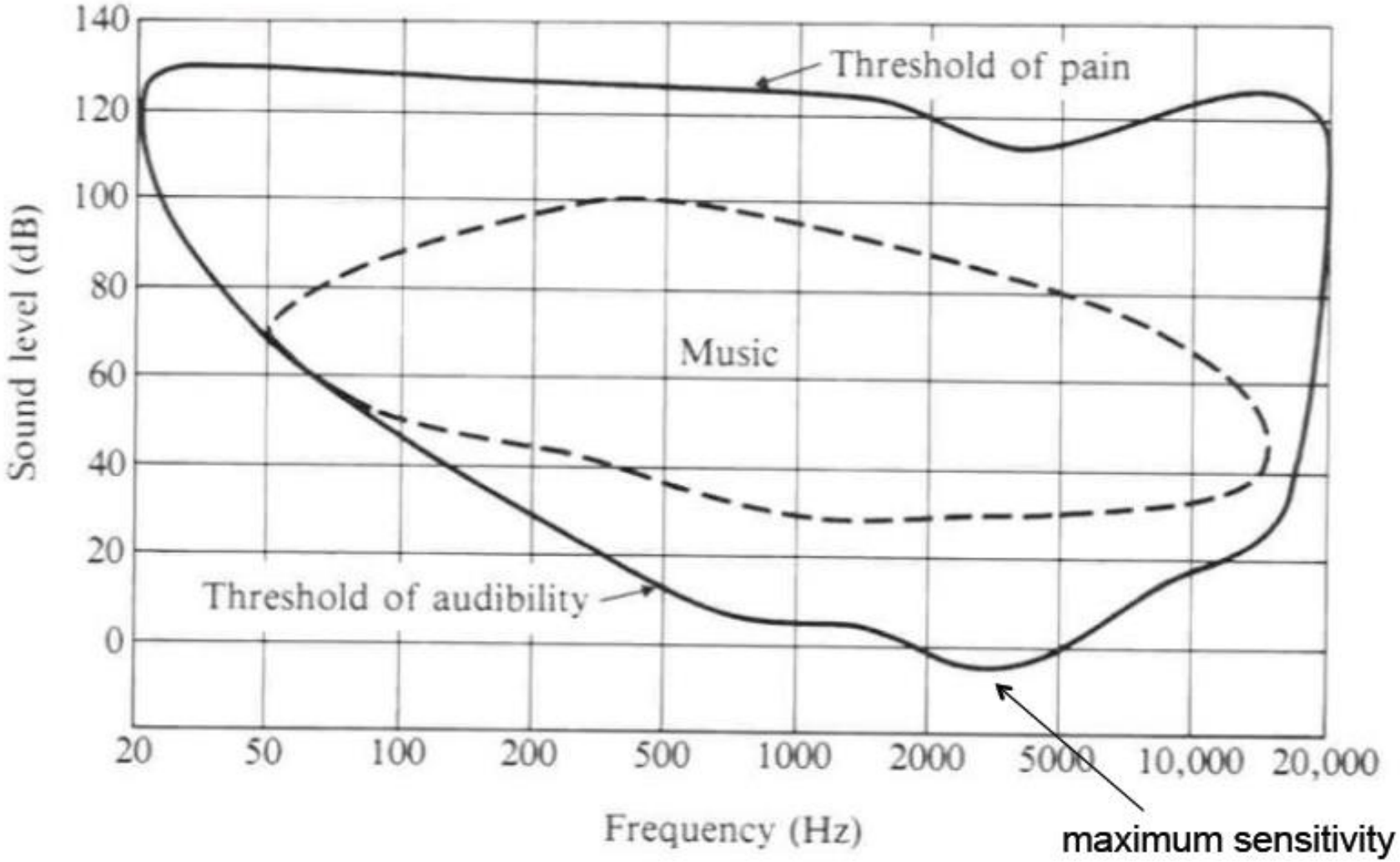


Sound Pressure Levels (dB)

SPL (dB)	Sound Source
160	Jet Engine — close up
150	Firecracker; Artillery Fire
140	Rock Singer Screaming into Microphone: Jet Takeoff
130	***Threshold of Pain***; .22 Caliber Rifle
120	Planes on Airport Runway; Rock Concert; Thunder
110	Power Tools; Shouting in Ear
100	Subway Trains; Garbage Truck
90	Heavy Truck Traffic; Lawn Mower
80	Home Stereo — 1 foot; Blow Dryer

SPL (dB)	Sound Source
70	Busy Street; Noisy Restaurant
60	Conversational Speech — 1 foot
50	Average Office Noise; Light Traffic; Rainfall
40	Quiet Conversation; Refrigerator; Library
30	Quiet Office; Whisper
20	Quiet Living Room; Rustling Leaves
10	Quiet Recording Studio; Breathing
0	***Threshold of Hearing***

Range of human hearing



Auditory masking

- The Human Auditory System (HAS) has a limited detection ability when a stronger signal occurs near (in frequency and time) to a weaker signal. In many situations, the weaker signal is imperceptible even under ideal listening conditions.
- We can consider two kinds of masking:
  - **Frequency masking** (also known as simultaneous masking) appears when different tones fire the same cochlear filters along the basilar membrane, which are coded as the same in the auditory nerve
  - **Temporal masking**: appear when a signal saturates a particular cochlear filter, during which time it does not register anything

Auditory Masking

- The human auditory system is often modelled as a filter bank which is based on particular perceptual frequency scale.
- These filters are called ‘critical-band’ filters
- From the point of view of perception, critical bands can ne treated as single entities within the spectrum.
- Signal components within a given critical band can be masked by other components within the same critical band.
- This is called **intra-band** masking.

- In addition, sounds on one critical band can mask sounds in different critical bands.

- This is called **inter-band** masking.

- While the masking process is very complex and only partially understood, the basic concepts can be successfully used in audio compression systems, so that better compression is achieved.

- Many people have examined the human auditory system and have concluded that the ear is primarily a frequency analysis device and can be approximated by a bandpass filters (known as the **critical-band filters**).

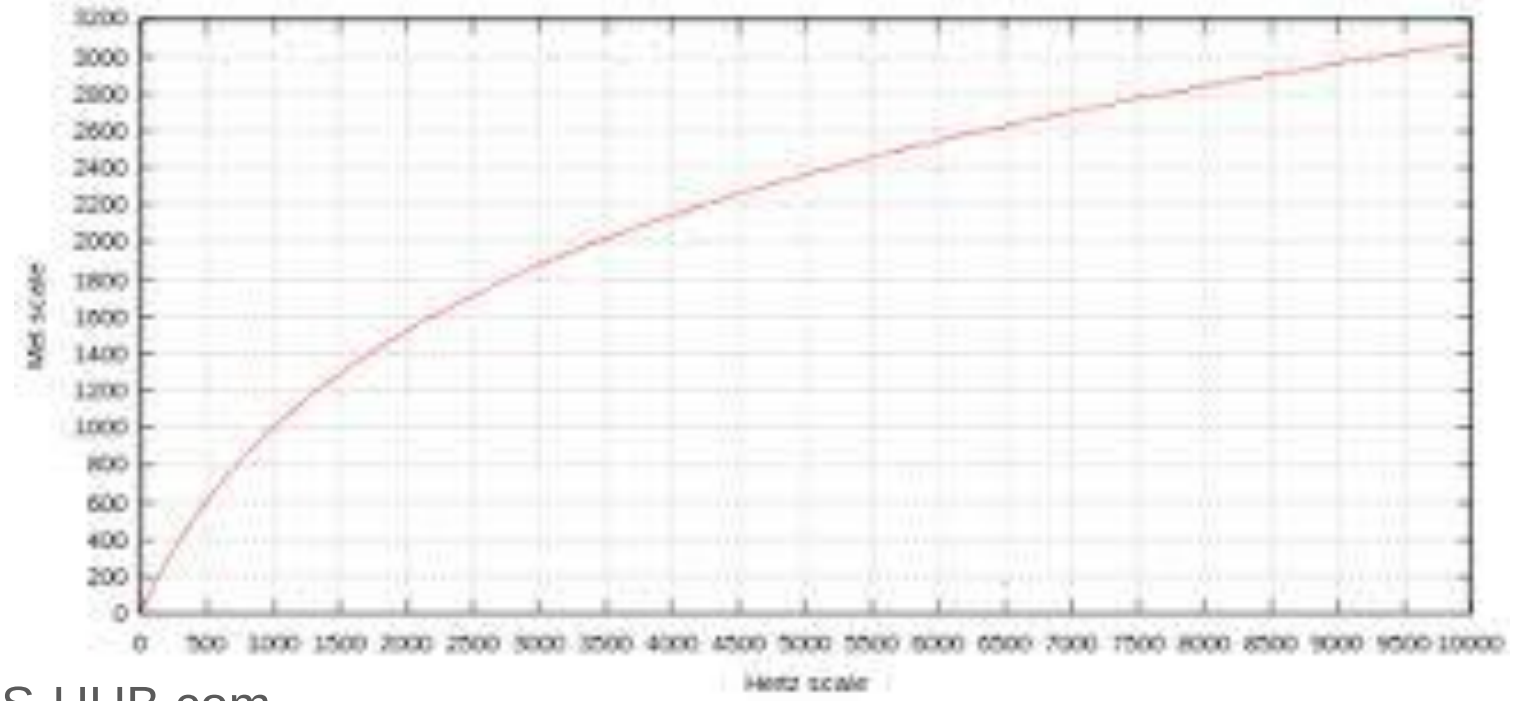
- Twenty five critical bands are required to cover frequencies of up to 20KHz.

- These filters may be spaced on a perceptual frequency scale known as ‘Bark scale’
- Experiments on the response of the basilar membrane in the ear have shown a relationship between acoustical frequency and perceptual frequency resolution
- A perceptual measure, called the Bark scale, provides the relationship between the two.
- The relationship between the frequency in Hz and the ‘critical band rate’ (with the unit of Bark) can be approximated by the following equations.

- Z_v (Bark) = $13.0\tan^{-1}(0.76f)$ f<1.5KHz

- Z_v (Bark) = $8.7 + 14.2\log_{10}(f)$, f>1.5KHz

- Where f is the frequency in KHz and Zv is the frequency in Barks. Figure below shows a plot of Barks vs. frequency (in KHz) up to 4 KHz



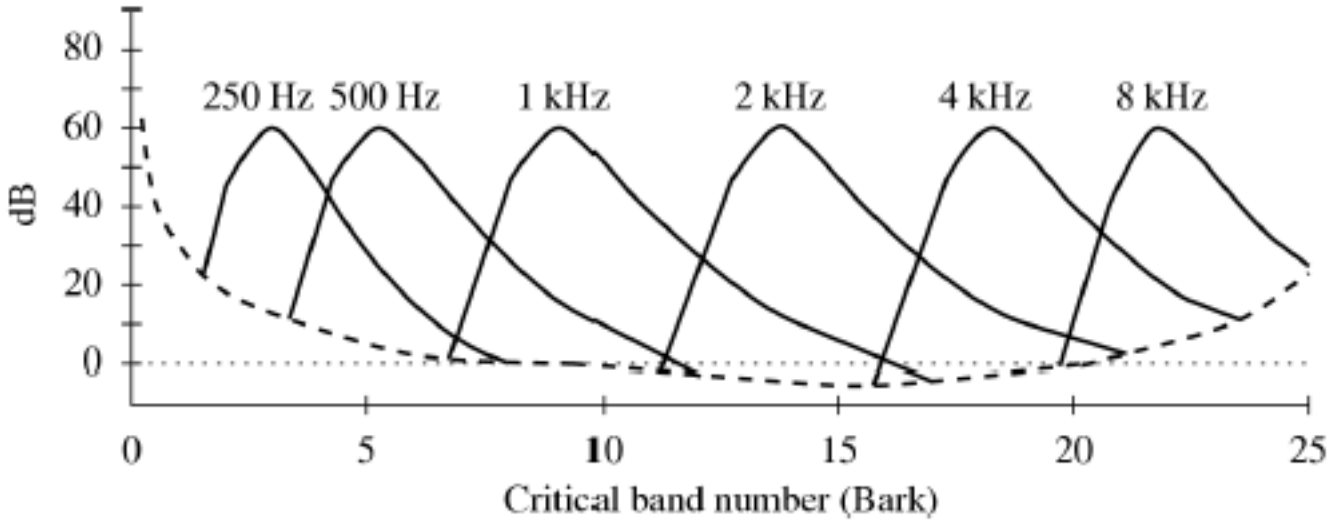
- Critical bandwidth is roughly constant at about 100 Hz for low frequency (<500Hz)

- For high frequencies, the critical bandwidth increases reaching approximately 700Hzat center frequencies around 4KHz.

- 25 critical bands are required to cover frequencies of up to 20KHz.

Bark Unit

- **Bark unit** is defined as the width of one critical band, for any masking frequency
- The idea of the Bark unit: every critical band width is roughly equal in terms of Barks (refer to Fig. 14.5)

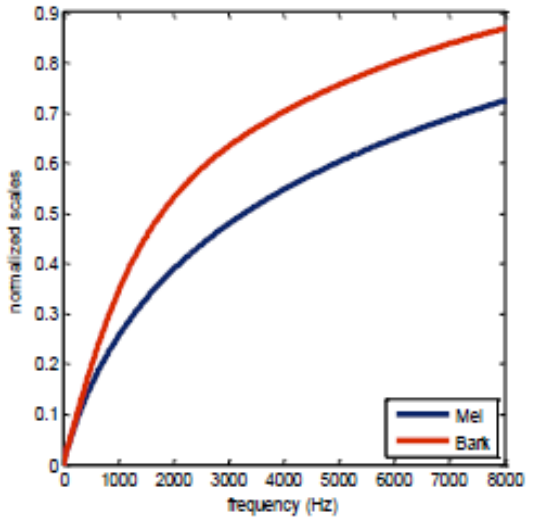


Effect of masking tones, expressed in Bark units

Two perceptual scales have been derived from critical bands

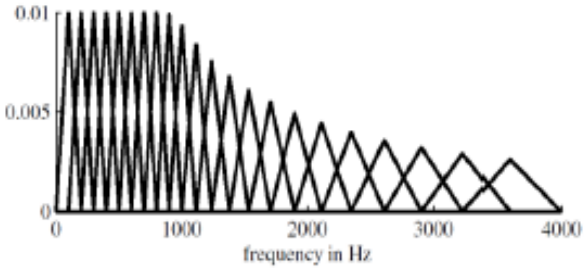
- Bark scale
  - Relates acoustic frequency to perceptual frequency resolution
  - One Bark equals one critical band

$$z = 13tan^{-1}\left(0.76\frac{f}{kHz}\right) + 3.5tan^{-1}\left(\frac{f}{7.5kHz}\right)$$



- Mel scale (more **Later on**)
  - Linear mapping up to $1\ kHz$, then logarithmic at higher frequencies

$$m = 2595log_{10}(1 + f/700)$$



[Rabiner & Schafer, 2007]

Critical Bands and Bandwidth

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
1	-	50	100	-
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

Masking

The effect of masking plays a very important role in hearing. It can be differentiated into two forms:

- Simultaneous masking (in Frequency)
- Non-simultaneous masking or Temporal masking (in time)

Simultaneous Masking

- An example of simultaneous masking would be the case where a person is having a conversation with another person where a loud track passes by. In this case, the conversation is severely disturbed and to continue the conversation successfully, the speaker has to raise his voice to produce more speech power and greater loudness.

- In music, similar effect take place when different instrument can mask each other and softer instrument become only audible when the loud instrument pauses.

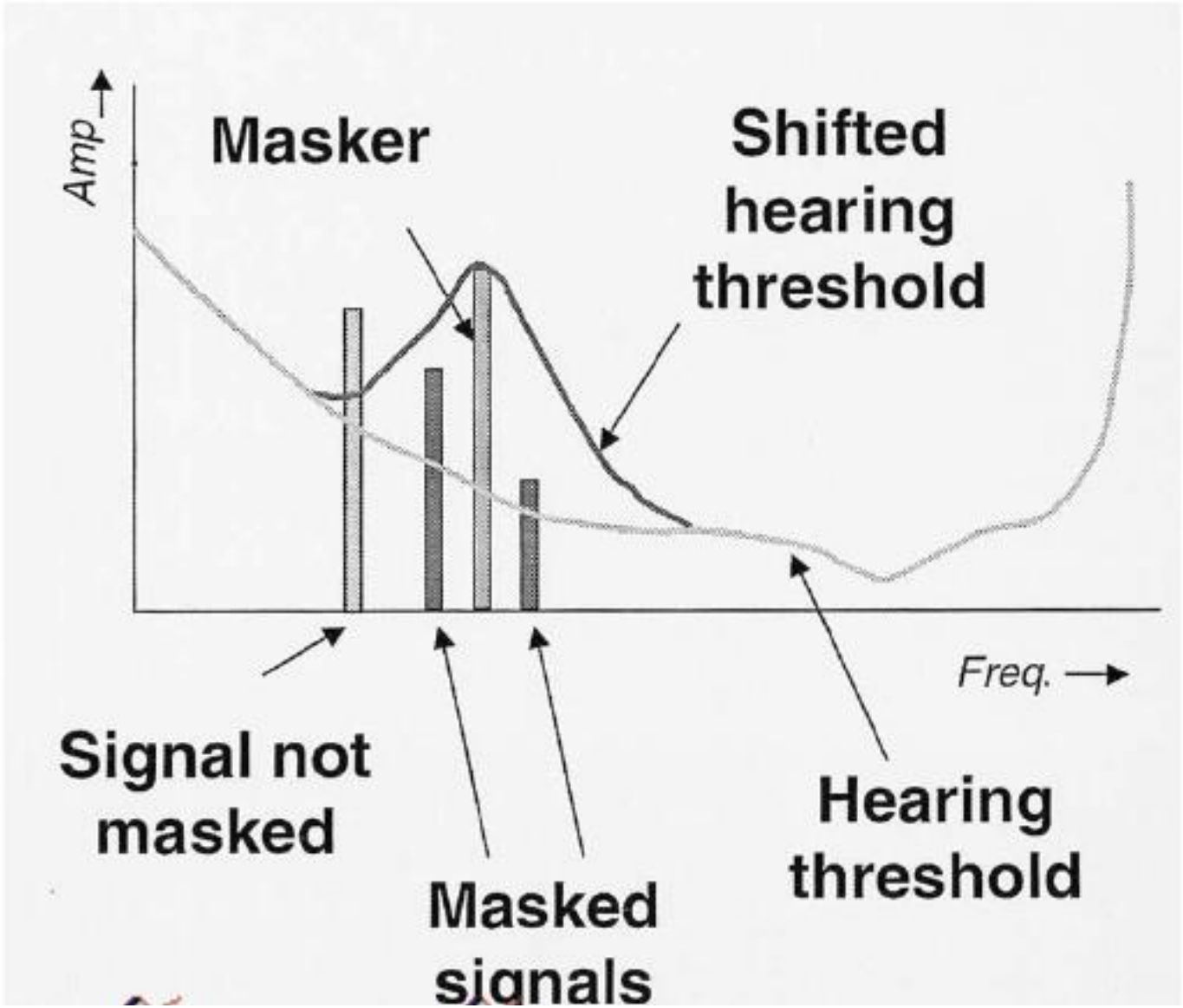
- Masking is usually described in terms of the minimum sound-pressure level of a test sound (a pure tone in most cases) that is audible in the presence of a masker.

- Most often, narrow-band noise of a given centre frequency and bandwidth is used as a masker.

- The excitation level of each masker is 60 dB.

- Comparing the results produced for different centre frequencies of the masker, we find the shapes of the masking curves are rather dissimilar irrespective of the frequency scaling (linear/log) used.

Frequency/simultaneous masking



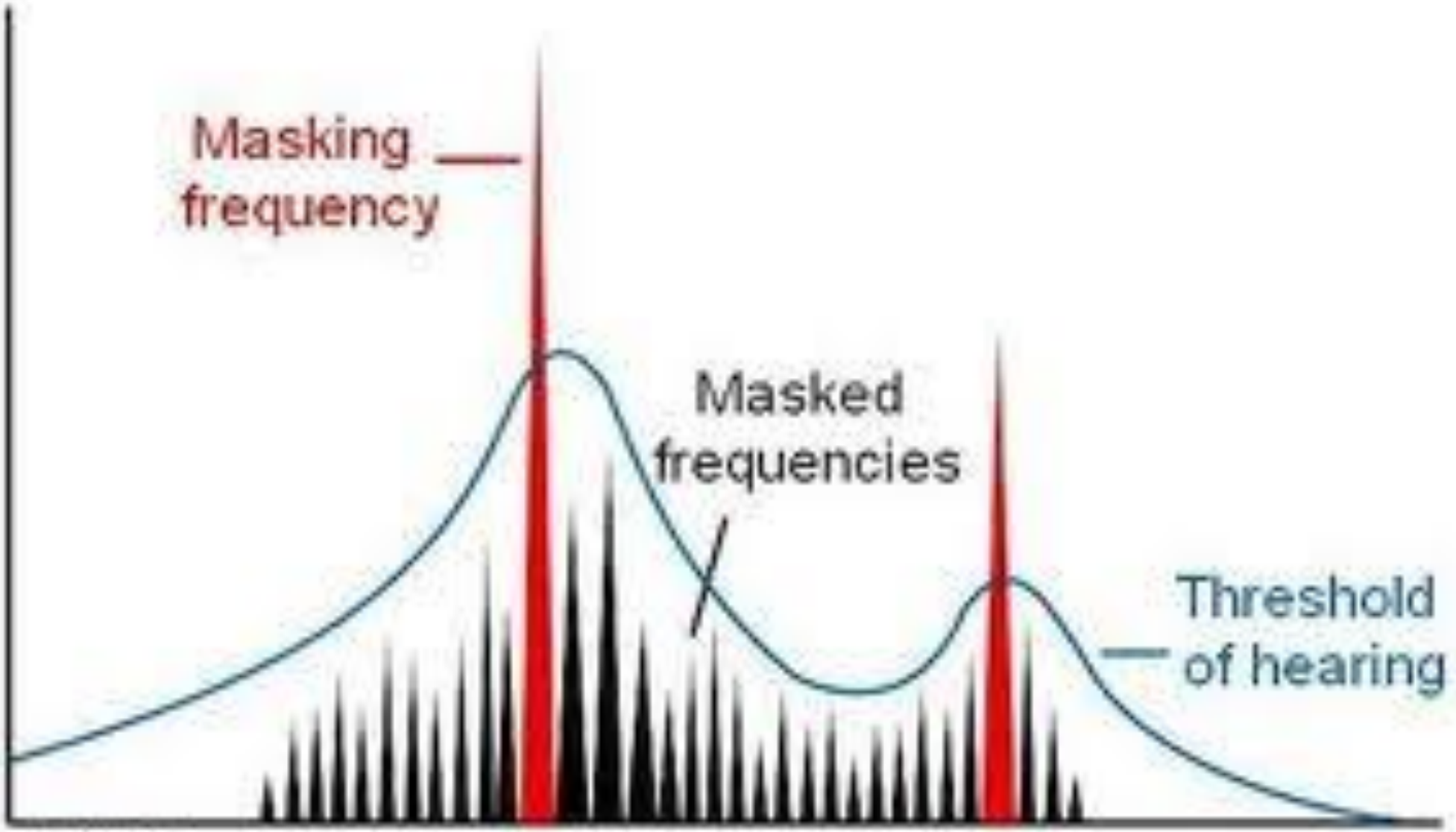


FIG. 1

Gammatone Filters

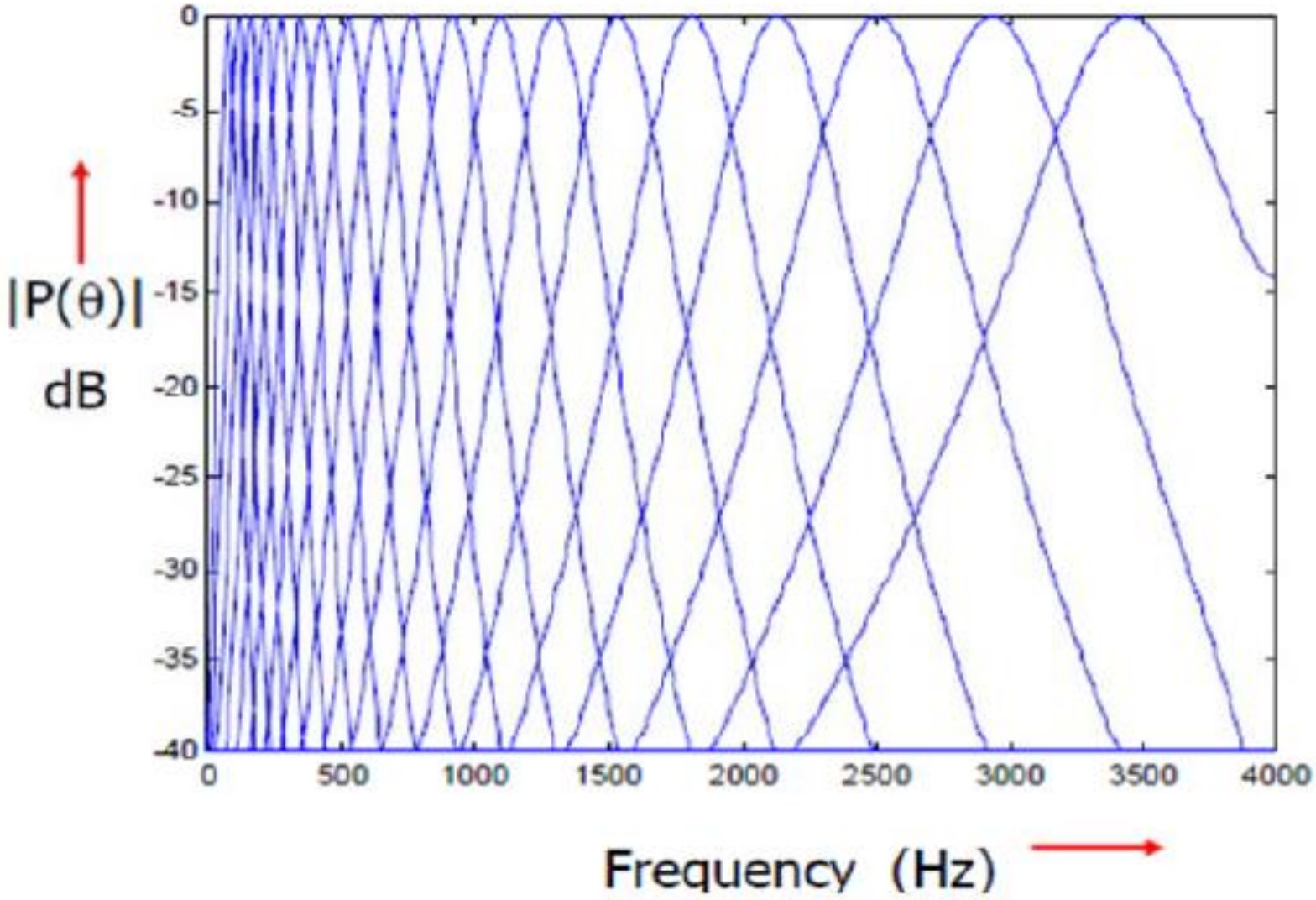
- Auditory filters are sometimes modelled by gammatone filters.
- Gammatone filters can be implemented as FIR or IIR filters.
- The impulse response of a Gammatone filter is given as

$$g(t) = at^{n-1}e^{-2\pi bERB(f_c)t}\cos(2\pi f_c t + \phi)$$

where, $ERB(f_c)$ is the bandwidth, f_c the centre frequency, a=1, b = 1.019 and n is the order of the filter.

$$ERB(f_c) = 24.7 + 0.108 f_c$$

Gammatone Filters



Nonsimultaneous masking

- Nonsimultaneous masking is also referred to as temporal masking. Temporal masking may occur when two sounds appear within a small interval of time.

- Two time domain phenomena play an important role in human auditory perception:

- Pre-masking

- Post-masking

- When the signal proceeds the masker in time, the condition is called post-masking; when the signal follows the masker in time, the condition is pre-masking,